Am J Cancer Res 2019;9(9):2047-2063
www.ajcr.us /ISSN:2156-6976/ajcr0098037

Original Article

The characterization of lung microbiome in lung cancer patients with different clinicopathology

Danhui Huang[1], Xiaofang Su[1], Man Yuan[1], Shujia Zhang[1], Jing He[1], Qiuhua Deng[1,2], Wenjun Qiu[3], Hangming Dong[1], Shaoxi Cai[1]

[1]*Chronic Airways Diseases Laboratory, Department of Respiratory and Critical Care Medicine, Nanfang Hospital, Southern Medical University, Guangzhou 510515, Guangdong, China;* [2]*Department of Respiratory and Critical Care Medicine, Zengcheng Branch of Nanfang Hospital, Southern Medical University, Zengcheng 511338, Guangdong, China;* [3]*First School of Clinical Medicine, Southern Medical University, 1838 North Guangzhou Avenue, Guangzhou 510515, Guangdong, China*

Received June 5, 2019; Accepted August 13, 2019; Epub September 1, 2019; Published September 15, 2019

Abstract: There were few knowledge concerned correlation between lung microbiome and different clinicopathology of lung cancer. Bronchial washing fluid (BWF) and sputum are commonly used sample types but there was no study comparing difference of microbiome between these two in lung cancer. In this study, we aimed to compare difference of microbiome between these two sample types and characterize lung microbiome in squamous cell lung carcinoma with (SCC_M1) or without distant metastasis (SCC_M0) and lung adenocarcinoma with (AD_M1) or without distant metastasis (AD_M0). We collected 40 BWF samples and 52 sputum samples from newly diagnosed lung cancer patients. Bacterial species were sequenced via 16S rRNA sequencing. Phylum *Proteobacteria* in BWF samples were significantly higher than sputum samples (Wilcoxon test, P = 0.003). At phylum level, microbiome of BWF samples was more similar to that of lung cancer tissues reported in the previous literature. LEFse analysis showed that in BWF group, genera *Veillonell*, *Megasphaera, Actinomyces* and *Arthrobacter* in AD_M0 were significantly higher than those in SCC_M0, and genera *Capnocytophaga* and *Rothia* in AD_M1 were significantly lower than that in SCC_M1. Compared with AD_M0, genus *Streptococcus* of AD_M1 was significantly lower, and genera *Veillonella* and *Rothia* in SCC_M1 were significantly higher than that in SCC_M1. Our study suggested that BWF samples might better reflect the microbiome of lung cancer tissues. In different metastatic states of lung cancer, differential genera between squamous cell carcinoma and adenocarcinoma were different. And in different histologic types of lung cancer, distant metastasis-related genera were not the same.

Keywords: Non-small cell lung cancer, microbiome, 16S rRNA, histology, metastasis

Introduction

Lung cancer is the most frequent cause of cancer death worldwide [1]. The five-year relative survival rate for localized, regional and distant lung cancer was 56%, 29%, 5% respectively [1].

Pathogens play an important role in carcinogenesis. *Helicobacter pylori* has been proved to promote the development of gastric cancer through inflammation and epithelial cell injury [2]. There were also some pathogens reported to be associated with lung cancer development. Previous studies suggested that a history of tuberculosis [3], Chlamydia pneumonia [4] and pneumococcal pneumonia [5] were associated with increased lung cancer risk. The lung has long been considered as a sterile space since Hilty M et al. firstly identified that microbiome existed in the lung of healthy people using bacterial 16S rRNA sequencing [6]. Since then, emerging evidences had suggested the link between lung microbiome and chronic lung disease, such as asthma, chronic obstructive pulmonary disease (COPD) [7].

Studies targeting at the association of microbiome and lung cancer remain in its early stage. Recent studies using next generation sequence have identified the lung microbiome of lung cancer patients was different from healthy people [8-15]. Most studies suggested that α diversity

[11-15] and β diversity [9, 10, 12-14] of lung microbiome of lung cancer were significant different from healthy controls. Besides, significant alteration of specific genera among lung cancer patients had been identified [8-10, 12-15].

Very few studies had investigated the association between lung microbiome and different clinicopathology of lung cancer. It was suggested that lung microbiome was associated with histologic classcification of lung cancer. By analyzing 165 normal adjacent tumor tissues of lung cancer (only 7 of them were in stage IV), a study found that genus *Thermus* in lung adenocarcinoma was more abundant than that in lung squamous cell carcinoma, while genus *Ralstonia* was lower [11]. K. Leigh Greathouse et al. analyzed 143 lung cancer tissue (only 3 of them were in stage IV) and found that genera *Acidovorax, Klebsiella, Rhodoferax* and *Anaerococcus* were more enriched in lung squamous cell than that in lung adenocarcinoma [13]. Few studies showed link between lung bacteria and distant metastasis of lung cancer. By analyzing 7 stage IV lung cancer and 151 I-IIIA stage lung cancer, genus *thermus* was found to be significant higher in stage IV lung cancer [11]. A study showed that Streptococcus up-regulated the expression of IL-6 through Toll-like receptor 2, which enhanced the matrix adhesion of non-small cell lung cancer cells and increased hepatic metastasis [16].

Lung cancer is a heterogeneous disease. Squamous cell carcinoma and adenocarcinoma are the two most common pathological types of lung cancer, which are characterized by different biological patterns, molecular biology and treatment strategies [17-19]. Lung squamous cell carcinoma and lung adenocarcinoma had different distant metastasis mechanisms [20-22]. Besides, different gene expression models and molecular biology existed in different stage of non-small cell lung cancer [22, 23]. Thus we speculated that histologic-related microbiome difference should be analyzed separately according to metastatic state of lung cancer and distant metastatic-related microbiome difference should analyzed separately according to pathological type.

Currently, several sample types (sputum, BWF, bronchial brushing tissue, surgical resection tissue) were used to study lung microbiome of patients with lung cancer, but heterogeneity between different samples was large. A study comparing the microbiome of spontaneous sputa and transplanted lung tissues in patients with cystic fibrosis showed that the relative abundance of the dominant genera in sputum were similar to that of lung tissue, but rare genera were significantly different [24]. Another study compared microbiome difference between induced sputum, bronchial aspirate, bronohoalveolar lavage fluid and bronchial brushing samples in patients with COPD [25]. The results showed that the taxonomy structure of induced sputum was similar to that of bronchial aspirate and the taxonomy structure of BWF was similar to that of bronchial brushing samples [25]. Tumor tissue is the ideal sample type for investigating microbiome of lung cancer, but most patients with advanced lung cancer are not indicted for surgery. The common alternative samples are sputum and BWF samples, but no current studies had compared the difference between these two sample types. Therefore, it is necessary to explore whether sputum and BWF can approximately reflect taxonomy structure of lung cancer tissue.

There are two purposes of this study. Firstly, we want to characterize lung microbiome in squamous cell lung carcinoma with (SCC_M1) or without distant metastasis (SCC_M0) and lung adenocarcinoma with (AD_M1) or without distant metastasis (AD_M0). Secondly, we want to compare the difference of microbiome between spontaneous sputum and BWF samples of lung cancer patients.

Materials and methods

Patient recruitment and samples collection

The study was approved by the Ethics Committee of Nanfang Hospital, Southern Medical University (NFEC-2018-158). Lung cancer patients were prospectively admitted in this study at NanFang Hospital, Southern Medical University. Patients were divided into 2 groups: BWF group and sputum group. In BWF group, 40 BWF samples from 40 newly diagnosed lung cancer patients were collected from October 2016 till September 2017. While in sputum group, 52 spontaneous sputum samples from 52 newly diagnosed lung cancer patients were collected from November 2017 till March 2018.

The inclusion criteria were as follows: pathologically diagnosed of lung squamous cell carcinoma or lung adenocarcinoma; aged 31-79 years old; did not receive surgery operation, radiotherapy or systemic therapy before sample collection; for sputum group, distant metastasis was confirmed by imaging examination or follow up visit; without community acquired pneumonia, acute bronchitis, acute exacerbation of chronic obstructive pulmonary disease bronchiectasis with infection or asthma; without prior history of other malignant diseases; had no fever or purulent or gray sputum.

The collection of BWF samples was assisted by member staffs in Bronchoscopy Room. Before the bronchoscopy, patients received a topical anaesthesia (lidocaine) by nebulizer and then were sedated with midazolam and fentanyl. The bronchoscope was wedged into patients' nasal cavity and then moved into lung. BW was performed following a standardized protocol. When the bronchoscope reached the "involved" airway containing the lung mass or the lung nodule, the bronchi were washed with 30-50 ml sterile 0.9% saline and approximately 15 ml BWF samples was acquired from each patient for further sequencing analysis. BWF samples were immediately stored at -20°C and transferred to the -80°C refrigerator within 1 week. When sterile saline was injected through a bronchoscope to the lung of a patient, it may contained some bacteria sequences which mixed with the lung microbiome. To evaluate contamination, 20 ml sterile 0.9% saline was injected via suctioning channel of bronchooscope and were then collected as negative controls. Spontaneous sputum was collected at the first day of hospitalization. Before sputum collection, patients were asked to rinse their mouth. Sputa were transferred into -20°C refrigerators within 3 hours and then transferred into -80°C within 1 weeks.

DNA extraction, 16S rRNA amplification, 16S rRNA sequencing

BWF samples and sputum samples were kept on dry ice and transferred to Sagene Biotechnology Company, GuangZhou. DNA was extracted from each sample using Hipure Bacterial DNA kit (Mageon, China) based on the manufacturer's recommendation. The V3-V4 region of 16S rRNA gene was amplified using specific primers (16S_341F: 5'-CCTAYGGGRBGCASCAG-3'; 16S_806R: 5-GGACTACNNGGGTATCTAAT). PrimeSTAR HS DNA Polymerase was used during PCR reaction. The PCR reaction procedure are as follows: initialization at 94°C for 5 min, followed by 31 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 30 s, and elongation at 72°C for 45 s, and a final elongation step at 72°C for 10 min. The length and concentration of the PCR products were detected by 1% agarose gel electrophoresis. Samples with a bright main strip were used for further experiments. Sequencing libraries were generated using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® sequencing (New England Biolabs, United States) following the manufacturer's recommendations. The library quality was assessed on a Qubit@ 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. Sequencing was conducted to generate 250-bp paired-end reads using an Illumina HiSeq 2500 sequencer according to the manufacturer's instructions.

Sequence data analysis

Raw data obtained after sequencing was further filtered with the following inclusion criteria: (i) reads containing less than 10% of unknown nucleotides. (ii) reads containing more than 80% of bases with quality. The filtered reads were then assembled into tags using FLASH (V1.2.11), according to overlap between paired-end reads with more than 10 bp overlap, and less than 2% mismatch. The software Mothur (V.1.34.0) was used to remove the redundant tags to get unique tags.

The software RDP classifier was used to classify tags into different taxonomies against SILVA database with Confidence Threshold of 0.5 [26]. USEARCH software (V8.0.1517) was used to cluster tags of more than 97% identity into operational taxonomic units (OTUs) [27]. The alpha diversity was evaluated by chao1 value, Shannon index and Simpson index. The OTU and Simpson rarefaction curve was used to evaluate whether the sequencing data amount was enough to cover all of the samples species and to reflect the species richness in samples. The β diversity was estimated using unweighted UniFrac or Bray Curtis distance and visualized by principal coordinate analysis (PCoA). Alpha diversity estimators and beta-diversity metrics were computed in online microbiome data analyse platform (MicrobiomeAnalyst) (https://www.microbiomeanalyst.ca/MicrobiomeAnalyst). Differential taxonomy was

Table 1. Demographics and clinical characteristics of BWF group and sputum group

Characteristics	BWF group	Sputum group
N	40	52
Age-mean (SD)	58.94 (9.04)	58.63 (9.27)
Gender		
Male, n (%)	33 (82%)	30 (70%)
Female, n (%)	7 (18%)	22 (30%)
BMI (kg/m^2)-mean (SD)	21.95 (3.15)	21.95 (3.43)
Smoking Status		
Current or former Smoker, n (%)	28 (70%)	25 (48%)
Never smoker, n (%)	12 (30%)	27 (52%)
Antibiotics use within 1 month		
Yes, n (%)	23 (58%)	28 (54%)
No, n (%)	17 (42%)	24 (46%)
Pathological diagnosis		
Squamous cell carcinoma, n (%)	21 (53%)	15 (29%)
Adenocarcinoma, n (%)	19 (47%)	37 (71%)
Distant metastasis		
Yes, n (%)	19 (40%)	32 (62%)
No, n (%)	14 (30%)	20 (38%)
Unidentified, n (%)	7 (24%)	0

Data were presented as median (standard deviation) for continuous variables or n (%) for counts.

identified by LEFse (Linear discriminant analysis (LDA) effect size) analysis online (http://huttenhower.sph.harvard.edu/galaxy).

Statistic analysis

Continuous variables were compared between groups by Wilcoxon rank-sum test or independent t test, and categorical variables were analyzed using Chi-square test or Fisher's exact test. Data are shown as the median (standard deviation) for continuous variables and number (%) for categorical variables. *P*-value < 0.05 was considered statistically significant. Geiphi was used to generate the correlation network between each genus. For the differential genera obtained by LEFse analysis, we used receiver operating characteristic curve (ROC) to evaluate the diagnostic value. Statistical analyses were carried out using R 3.5.1.

Results

Characteristics of the subjects

The BWF group consisted of 40 non-small cell lung cancer patients. 21 were lung squamous cell carcinoma and 19 were lung adenocarcinoma. The sputum group consisted of 52 patients, including 15 patients with lung squamous cell carcinoma and 37 patients with lung adenocarcinoma. The basic information of the two groups was shown in **Table 1**. For BWF group, the average number of trimmed sequence reads of 40 patients was 40908 (16750, 57557). For sputum group, the average number of trimmed sequence reads number of 52 lung cancer patients was 34685 (20233, 44578). OTU and Simpson rarefaction curve was constructed to evaluate sequence depth (Supplementary Figure 1). The results indicated that sequence depth of BWF samples and sputum samples was sufficient enough to reach a reliable estimate of microbiome structure.

Taxonomy composition of BWF samples and sputum samples from lung cancer patients

5 negative controls were collected in order to evaluate contamination of BWF samples. Because of low DNA concentration, 1 negative control could not yield sufficient data for sequence analyse. Unweighted unifrac PCOA plot at the OTU level was constructed (Supplementary Figure 2) and the result showed that taxonomy structure between BWF samples from lung cancer patients and negative controls were significant different, which indicated that the microbiome structure of BWF samples had not been largely effected by contamination. Use K. Leigh Greathouse et al.'s [13] methods for reference, we firstly removed OTU sequences with relative abundance ≥ 5% in negative controls; then we deleted the remaining OTUs belonged to putative contaminant genera including *Halomonas*, *Ralstonia* and *Acinetobacter*. Genera with relative abudance ≥ 5% were considered as contaminant genera.

Phyla and genera that were ≥ 2% were considered as dominant. In BWF group, the dominant phyla were *Firmicutes* (32%), *Proteobacteria* (30%), *Bacteroidetes* (25%), *Actinobacteria* (6%) *Fusobacteria* (3%) and *Cyanobacteria* (2%). The dominant genera were *Prevotella*

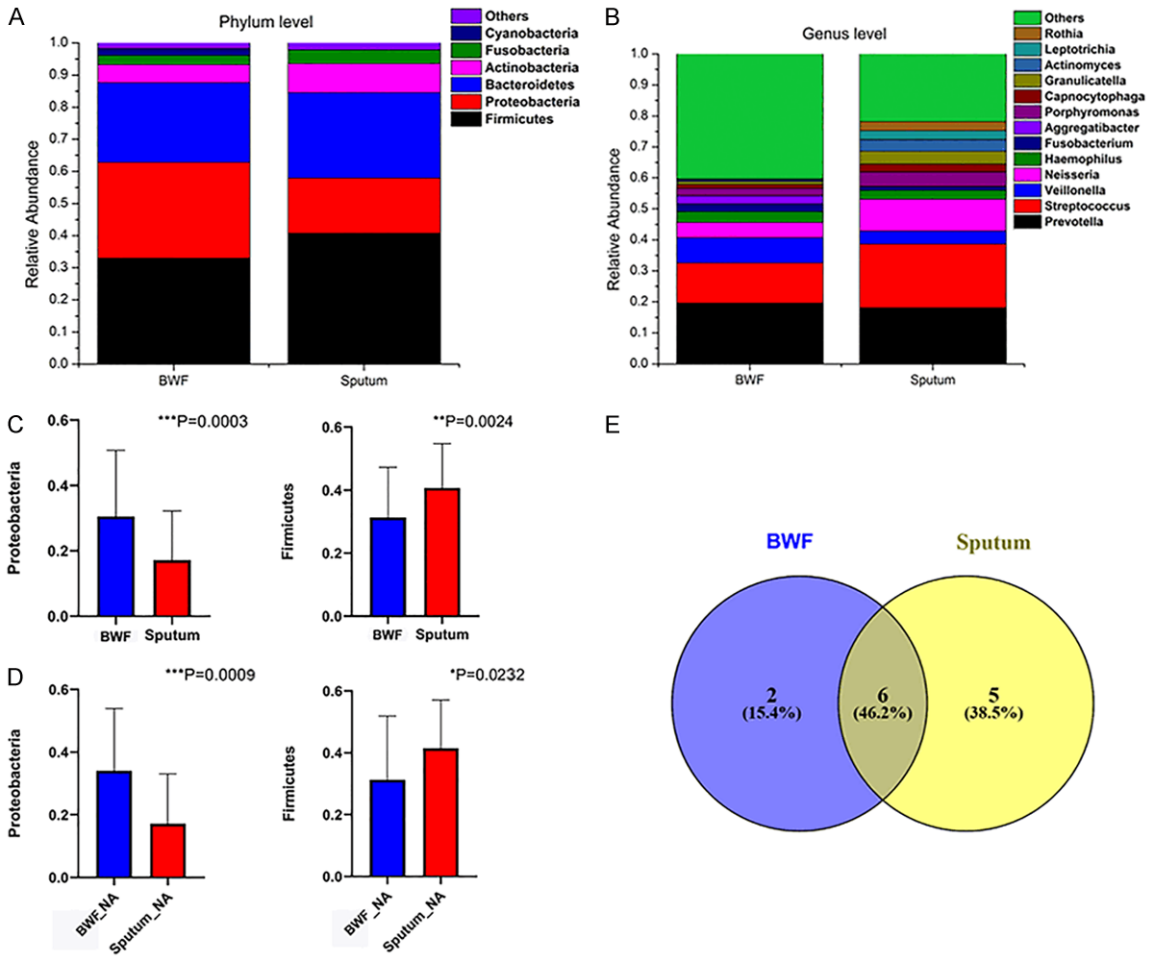


Figure 1. Taxonomic composition of microbiome community of BWF group and sputum group. A. Dominant phyla of BWF group and sputum group; B. Dominant genera of BWF group and sputum group; C. Differentially abundant of phylum *Proteobacteria* and *Firmicutes* between BWF group and sputum group; D. Antibiotics subgroup analysis of differentially abundant of phylum *Proteobacteria* and *Firmicutes* between BWF group and sputum group; E. Venn graph shows the number of collective and exclusive dominant genera between BWF and sputum samples of lung cancer patients. *P* values were calculated with Wilcoxon test.

(20%), *Streptococcus* (12%), *Veillonella* (8%), *Neisseria* (5%), *Haemophilus* (3%), *Fusobacterium* (3%), *Aggregatibacter* (2%) and *Porphyromonas* (2%). While in sputum group, the dominant phyla were *Firmicutes* (43%), *Bacteroidetes* (27%), *Proteobacteria* (17%), *Actinobacteria* (8%) and *Fusobacteria* (3%). The dominant genera were *Streptococcus* (23%), *Prevotella* (20%), *Neisseria* (10%), *Porphyromonas* (5%), *Veillonella* (5%), *Granulicatella* (5%), *Actinomyces*(3%), *Rothia* (3%), *Haemophilus* (3%), *Leptotrichia* (3%), *Capnocytophaga* (2%), *Atopobium* (2%), *Lautropia* (2%). **Figure 1A**, **1B** listed the dominant phyla and genera of BWF group and sputum group.

Phylum *Proteobacteria* in BWF group were significantly higher than that in sputum group (Wilcoxon test, P = 0.0003), while phylum *Firmicutes* were significantly decreased (Wilcoxon test, P = 0.0024) (**Figure 1C**). In subgroup analysis of lung cancer patients without antibiotics use within 1 month, *Proteobacteria* in BWF group were also significantly higher (Wilcoxon test, P = 0.0009), while phylum *Firmicutes* were also significantly decreased (Wilcoxon test, P = 0.0232) (**Figure 1D**). To date, the largest study of lung cancer microbiome which included 143 lung cancer tissues showed that the main phylum of lung cancer was *Proteobacteria* with a relative abundance of 70% [13]. Another microbiome study with 165 adjacent normal lung cancer tissue samples showed that the main phylum was also *Proteobacteria*, with a relative abundance of 60% [11]. Therefore, according to these 2 previ-

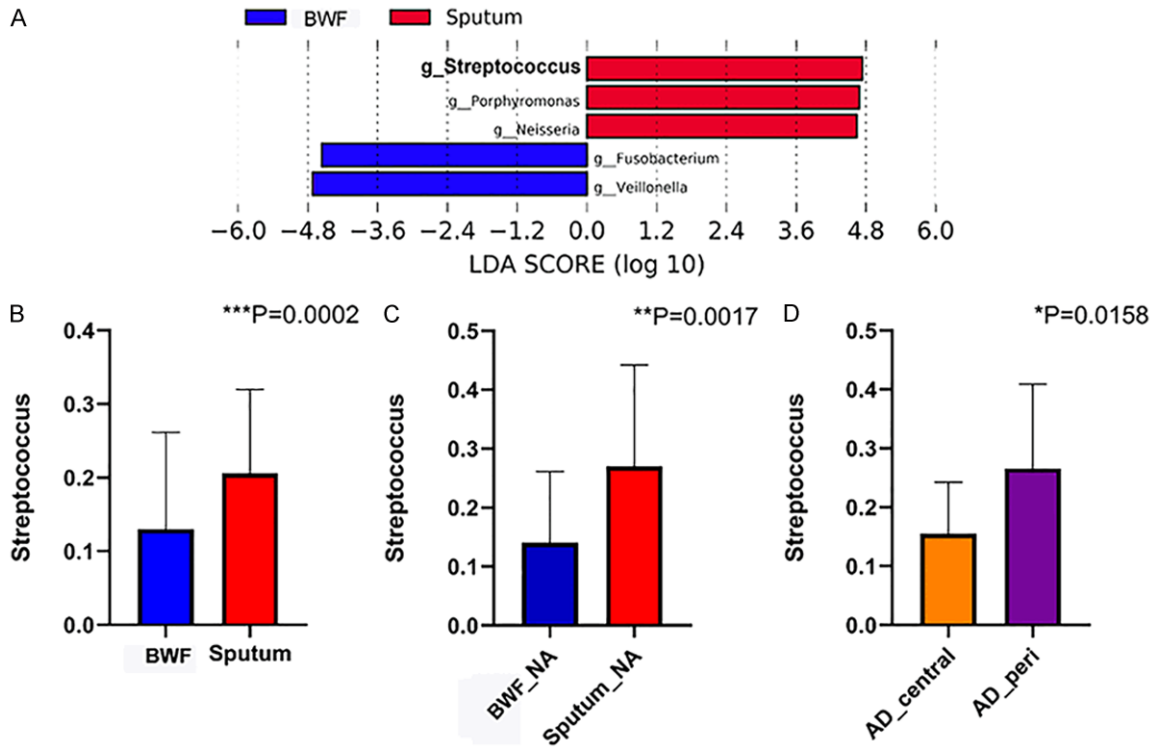


Figure 2. Differentially abundant genera among BWF group and sputum group, and between AD_central and AD_peri. A. LEFse analysis of collective 6 dominant genera between sputum and BWF group; B. Differentially abundant of genus *Streptococcus* between BWF group and sputum group; C. Antibiotics subgroup analysis of differentially abundant of genus *Streptococcus* between BWF group and sputum group; D. Differentially abundant of genus *Streptococcus* between AD_central and AD_peri in sputum group. *P* values were calculated with Wilcoxon test. AD_central: central lung adenocarcinoma in sputum group; AD_peri: peripheral lung adenocarcinoma.

ous researches [11, 13], at phylum level, we considered BWF samples might better reflect microbiome structure of lung cancer tissue.

Since sputum is nutrient-rich and some bacterium may easily grow under such condition, we speculate that bacteria breed may be one reason for the difference between BWF and sputum. Breed of genera with high relative abundance were more easily to be detected. 6 dominant genera were shared across sputum and BWF samples (**Figure 1E**), which includes *Streptococcus*, *Prevotella*, *Neisseria*, *Porphyromonas*, *Veillonella* and *Haemophilus.* LEFse was conducted to further evaluate the difference of these 6 genera between BWF samples and sputum samples. The result showed that *Streptococcus,* which belonged to *Firmicutes,* was significantly higher in sputum samples (Wilcoxon test, P = 0.0002) (**Figure 2A**, **2B**). It was also the genus that had the greatest influence on the distinction between sputum and BWF samples with LDA score = 4.8. In subgroup analysis of lung cancer patients without antibiotics use within 1 month, *Streptococcus* was also significantly increased in sputum group (Wilcoxon test, P = 0.0017) (**Figure 2C**). *Streptococcus* is the most common pathogen of community-acquired pneumonia in China [28] and easily grow in lungs under favourable environment applied by sputum. In addition, we further found that *Streptococcus* in peripheral lung adenocarcinoma was significantly higher than that of central lung adenocarcinoma in sputum group (Wilcoxon test, P = 0.0158) **Figure 2D**). Basic information included age (independent t test, P = 0.334), BMI index (independent t test, P = 0.372), sex (chi-square test, P = 0.235) and smoking status (chi-square test, P = 0.285) and use of antibiotics (P = 0.285) were comparable between 2 groups (Supplementary Table 1). In addition, there was no significant difference in *Streptococcus* between peripheral lung adenocarcinoma and central lung adenocarcinoma in BWF group (Wilcoxon test, P = 0.1774). We speculate that the difference between peripheral lung adenocarcinoma and central lung adenocarcinoma

Table 2. Demographics and clinical characteristics among 4 groups of BWF group

Characteristics	SCC_M0	AD_M0	SCC_M1	AD_M1	*P* value
N	7	7	7	12	
Age-mean (SD)	58.71 (8.62)	58.42 (12.23)	64.29 (7.34)	57.58 (9.85)	0.522[a]
Gender					0.152[b]
Male, n (%)	6 (86%)	6 (86%)	7 (100%)	7 (58%)	
Female, n (%)	1 (14%)	1 (14%)	0	5 (42%)	
BMI (kg/m^2)-mean (SD)	24.63 (3.15)	22.34 (2.48)	20.26 (3.37)	21.61 (3.12)	0.076[a]
Smoking Status					0.690[b]
Current or former Smoker, n (%)	5 (71%)	4 (57%)	6 (86%)	7 (58%)	
Never smoker, n (%)	2 (29%)	3 (43%)	1 (14%)	5 (42%)	
Antibiotics use within 1 month					0.864[b]
Yes, n (%)	4 (57%	3 (43%)	5 (71%)	6 (50%)	
No, n (%)	3 (43%	4 (57%)	2 (29%)	6 (50%)	

Data were presented as median (standard deviation) for continuous variables or n (%) for counts. a: One way ANOVA, b: Fisher exact test.

Table 3. Demographics and clinical characteristics among 4 groups of sputum group

Characteristics	SCC_M0	AD_M0	SCC_M1	AD_M1	*P* value
N	6	14	9	23	
Age-mean (SD)	64.839.93)	60.00 (8.81)	59.11 (7.27)	56.00 (9.6)	0.182[a]
Gender					0.538[b]
Male, n (%)	4	7	7	12	
Female, n (%)	2	7	2	11	
BMI (kg/m^2)-mean (SD)	21.76 (2.30)	21.99 (4.48)	22.00 (3.34)	21.95 (3.17)	0.999[a]
Smoking Status					0.322[b]
Current or former Smoker, n (%)	4	7	6	8	
Never smoker, n (%)	2	7	3	15	
Antibiotics use within 1 month					0.689[b]
Yes, n (%)	2	8	6	13	
No, n (%)	4	6	3	10	

Data were presented as median (standard deviation) for continuous variables or n (%) for counts. a: One way ANOVA, b: Fisher exact test.

may be attributed to the *Streptococcus* breed, because it takes more time for sputum to be removed from the lung to upper airway in peripheral lung adenocarcinoma than central lung adenocarcinoma. These suggested that the difference between spontaneous sputum and BWF samples might be partly attributed by bacteria breed.

Characterization of lung microbiome with different histologic types

Previous studies suggested that lung microbiome was related with distant metastasis of lung cancer [11, 29]. Therefore, we only selected lung cancer patients whose existence of distant metastasis could be confirmed by image examination or follow-up visit. In total, 33 lung cancer patients in BWF group and 52 lung cancer patients in sputum group were chosen for further analysis. Lung cancer patients in BWF group and sputum group were divided into 4 groups: lung squamous cell carcinoma without distant metastasis (SCC_M0), lung squamous cell carcinoma with distant metastasis (SCC_M1), lung adenocarcinoma without distant metastasis (AD_M0) and lung adenocarcinoma with distant metastasis (AD_M1). Age, sex, BMI index, smoking and antibiotic use were comparable between the four groups in both BWF group and sputum group (**Tables 2** and **3**).

Chao1, Simpson index and Shannon index were selected to estimate the α diversity of lung

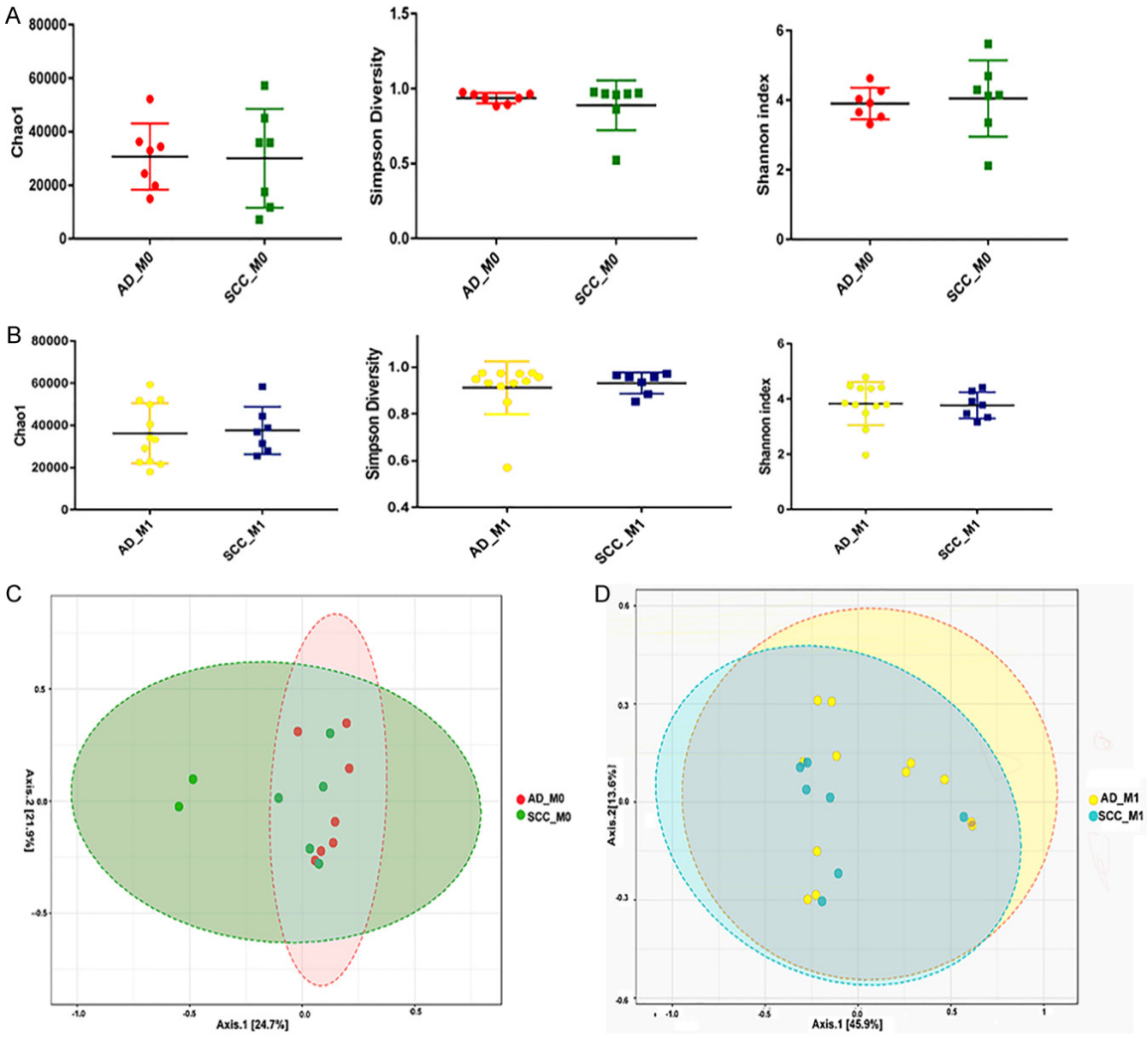


Figure 3. Comparison of α diversity and β diversity of BWF microbiome from different histologic types. A. α diversity between AD_M0 and SCC_M0; B. α diversity between AD_M1 and SCC_M1; C. PCOA plot based on Bray-Curtis distance of BWF microbiome between AD_M0 and SCC_M0; D. PCOA plot based on Bray-Curtis distance of BWF microbiome between AD_M1 and SCC_M1.

microbiome community. In BWF group, α diversity between SCC_M0 and AD_M0 was similar (P = 1.0 for chao1; P = 0.620 for Simpson index; P = 0.456 for Shannon) (**Figure 3A**). And for comparison between SCC_M1 and AD_M1, α diversity was also similar (P = 0.773 for chao1; P = 0.837 for Simpson index; P = 0.536 for Shannon) (**Figure 3B**). In the sputum group, we found that there was also no significant difference of α diversity between SCC_M0 and AD_M0, and between SCC_M1 and AD_M1 (Supplementary Figure 3A, 3B). Together, these results suggested that there was no significant difference in α diversity between lung squamous cell carcinoma and lung adenocarcinoma.

β diversity based on Bray Curtis distance was used to estimate the β diversity of lung taxonomy community structure in lung cancer patients. In BWF group, the results showed that taxonomy structure between SCC_M0 and AD_M0 (PERMONOVA test, P = 0.375) and taxonomy structure between SCC_M1 and AD_M1 (PERMONOVA test, P = 0.246) were similar (**Figure 3C**, **3D**). Similarly, in the sputum group, taxonomy structure between SCC_M0 and AD_M0 (PERMONOVA test, P = 0.386) and taxonomy structure between SCC_M1 and AD_M1 (PERMONOVA test, P = 0.829) were also similar (Supplementary Figure 3C, 3D). Together these results suggested that there was no significant difference in taxonomy structure

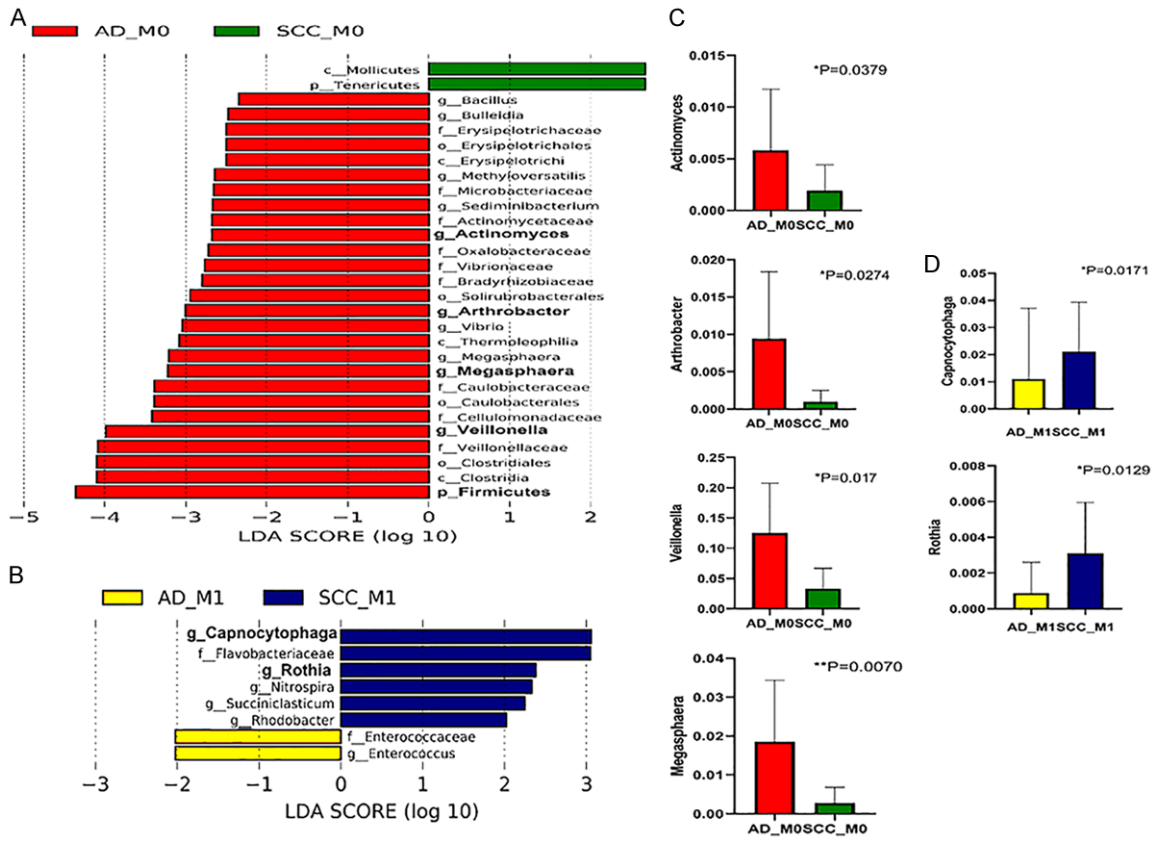


Figure 4. Differentially abundant taxonomy between different histologic types in BWF group. A. Differentially abundant taxonomy between SCC_M0 and AD_M0 identified by LEFse; B. Differentially abundant taxonomy between SCC_M1 and AD_M1 identified by LEFse; C. Differentially abundant of genera *Veillonella*, *Megasphaera*, *Actinomyces*, *Arthrobacter* between SCC_M0 and AD_M0; D. Differentially abundant of genera *Capnocytophaga* and *Rothia* between SCC_M1 and AD_M1. *P* values were calculated with Wilcoxon test. Phyla and genera that were ≥ 0.1% were shown as bold fonts in LEFse plots.

between lung squamous cell carcinoma and lung adenocarcinoma.

Considering that the abundance of some genera might be distorted by bacteria breed in sputum, we only selected BWF samples to analyze differential taxonomy between different histologic types. In stage I-III, for taxonomy ≥ 0.1%, LEFse analysis showed that compared with SCC_M0, phylum *Firmicutes* and genera *Veillonella, Megasphaera, Actinomyces, Arthrobacter* were significantly increased in AD_M0 patients (**Figure 4A**, **4C**; Supplementary Table 2). While in stage IV, compared with SCC_M1, genera *Capnocytophaga* and *Rothia* decreased significantly in AD_M1 (**Figure 4B**, **4D**; Supplementary Table 3). In order to investigate whether the difference of lung microbiome between different histologic types is related to smoking, we divided 33 BWF samples into 2 groups: current or former smoking group and non-smoking group. The result showed that the differential genera between SCC_M0 and AD_M0 and between SCC_M1 and AD_M1 was not related to smoking status (Wilcoxon test, P > 0.05, Supplementary Table 6). The results suggested that the differential genera related to histologic types should be separately analyzed according to different distant metastatic states.

Characterization of lung microbiome of different distant metastatic states

In BWF group, α diversity between SCC_M0 and SCC_M1 was similar (P = 0.456 for chao1; P = 0.532 for Simpson index; P = 0.383 for Shannon) (**Figure 5A**). And for the comparison between AD_M0 and AD_M1, α diversity was also similar (P = 0.837 for chao1; P = 0.773 for Simpson index; P = 0.773 for Shannon) (**Figure 5B**). In sputum group, we found that there was also no significant difference of α

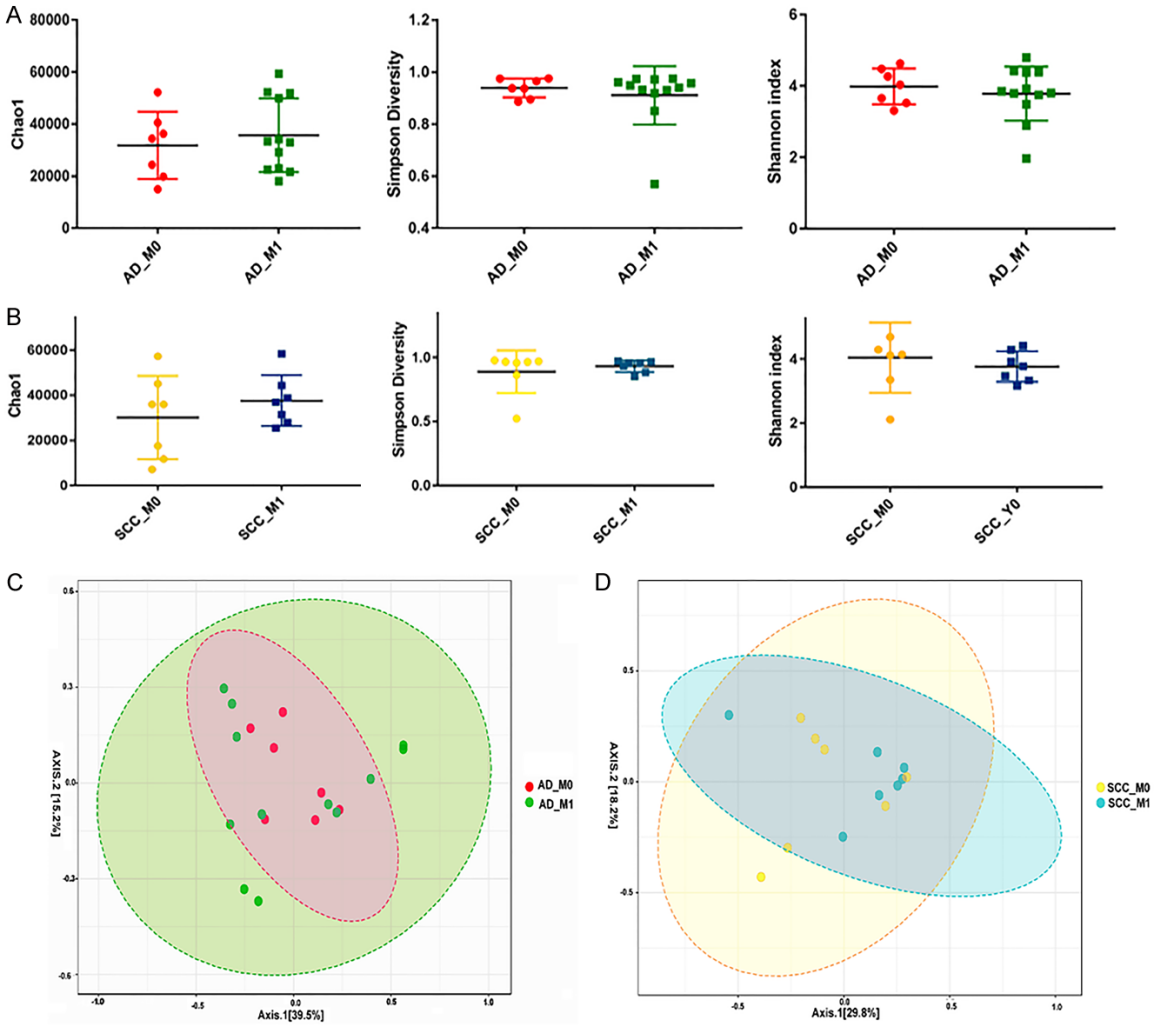


Figure 5. Comparison of α diversity and β diversity of BWF microbiome from different metastatic states. A. α diversity between AD_M0 and AD_M1; B. α diversity between SCC_M0 and SCC_M1; C. PCOA plot based on Bray-Curtis distance of BWF microbiome between AD_M0 and AD_M1; D. PCOA plot based on Bray-Curtis distance of BWF microbiome between SCC_M0 and SCC_M1.

diversity between SCC_M0 and SCC_M1, and between AD_M0 and AD_M1 (Supplementary Figure 4A, 4B). These results suggested that α diversity of lung microbiome was not related to the distant metastasis of lung cancer.

We used Bray Curtis distance to calculate β diversity of lung microbiome. The results showed that taxonomy structure between SCC_M0 and SCC_M1 (PERMONOVA test, P = 0.370) and taxonomy structure between AD_M0 and AD_M1 (PERMONOVA test, P = 0.693) were similar (**Figure 5C**, **5D**). Similarly, in the sputum group, SCC_M0 and SCC_M1 (PERMONOVA test, P = 0.788), AD_M0 and AD_M1 (PERMONOVA test, P = 0.776) also had similar taxonomy structure (Supplementary Figure 4C, 4D).

In BWF group, for taxonomy ≥ 0.1%, LEFse analysis showed a significant decrease in phylum *Firmicutes* and genus *Streptococcus* in AD_M1 compared with AD_M0 (**Figure 6A**, **6B**; Supplementary Table 4). ROC analysis showed *Streptococcus* had a moderate value in predicting distant metastasis of lung adenocarcinoma (AUC 0.787, 95% CI 0.546-1.0) (**Figure 6C**). In patients with lung squamous cell carcinoma, genera *Veillonella* and *Rothia* in SCC_M1 was significantly higher than that in SCC_M0 (**Figure 6D**, **6E**; Supplementary Table 5). ROC analysis showed that *Veillonella* (AUC 0.898, 95% CI 0.732-1.0) and *Rothia* (AUC 0.877, 95% CI 0.689-1.0) had a moderate diagnostic value in differentiating SCC_M1 from SCC_M0 (**Figure 6F**). The results suggested that the differential

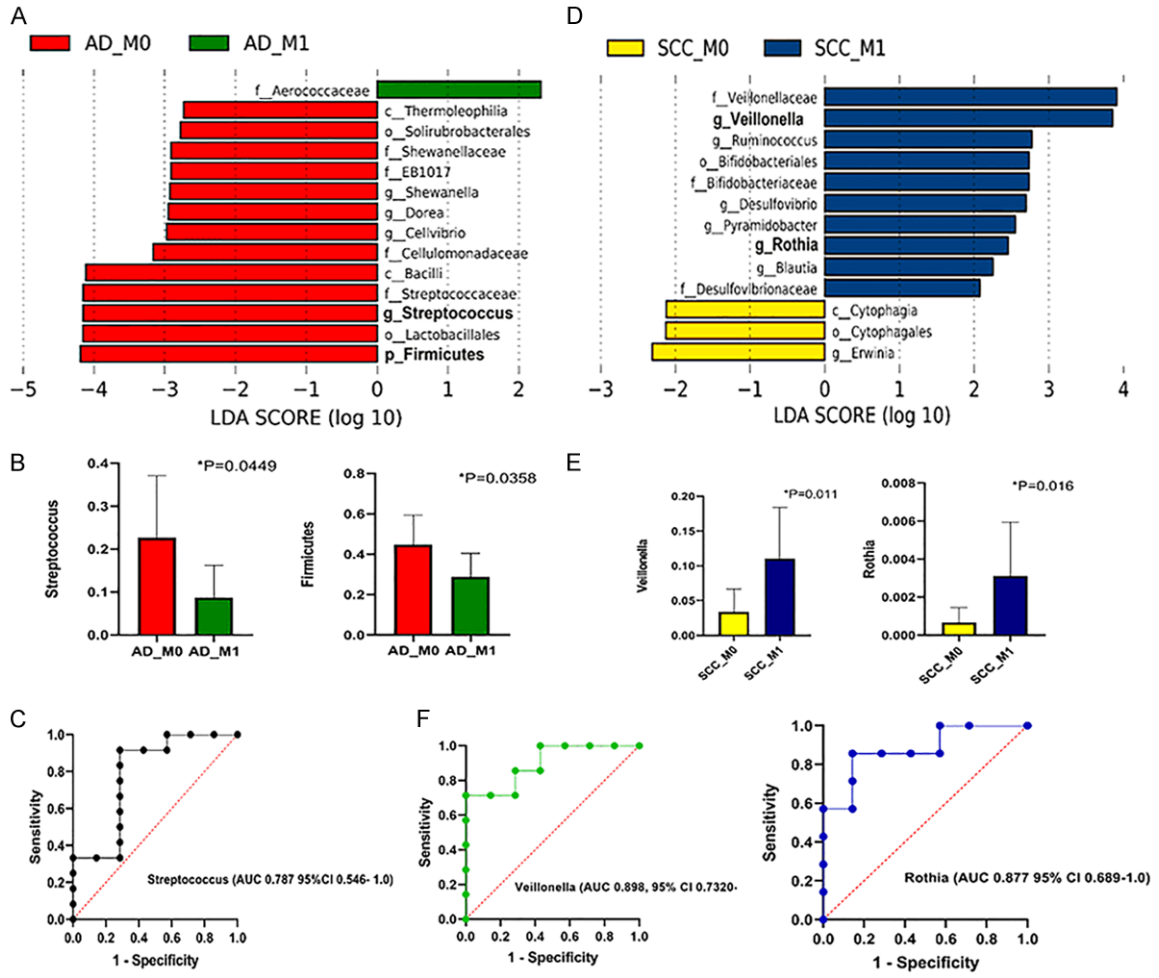


Figure 6. Differentially abundant taxonmoy between different metastatic states in BWF group. A. Differentially abundant taxonomy between AD_M0 and AD_M1 identified by LEFse; B. Differentially abundant of genus *Streptococcus* and phylum *Firmicutes* between AD_M0 and AD_M1; C. ROC curve with genus Streptococcus to predict distant metastasis of lung adenocarcinoma; D. Differentially abundant taxonomy between SCC_M0 and SCC_M1 identified by LEFse; E. Differentially abundant of genera Veillonella and Rothia between SCC_M0 and SCC_M1; F. ROC curve with genus Veillonella and Rothia to predict distant metastasis of squamous cell carcinoma. *P* values were calculated with Wilcoxon test. Phyla and genera that were ≥ 0.1% were shown as bold fonts in LEFse plots.

genera related to distant metastasis should be separately analyzed according to histologic types.

Discussion

Comparison of the microbiome among BWF samples and sputum samples from lung cancer patients

In this study, we compared the difference between BWF samples and spontaneous sputum samples of lung cancer patients, in order to determine whether BWF samples and sputum samples are suitable for representing the microbiome of lung cancer tissue. The most predominant phyla in BWF samples were *Firmicutes* and *Proteobacteria* and the most common genera were *Prevotella* (18%). While in sputum samples, *Firmicutes* (43%) was the most predominant phylum and *Streptococcus* (23%) was the most common genus. *Proteobacteria* in the sputum samples was significantly decreased compared with the BWF samples. H. Dean Hosgood III et al. collected sputum samples from 8 lung cancer patients and reported that the most abundant phylum in sputum was *Firmicutes*, and the most abundant genus was *Streptococcus*, which were similar to our findings [12]. In our study, although we did not collect lung cancer tissue

samples for 16S rRNA sequencing, but compared with the two largest studies [11, 13] to date which evaluate the microbiome of lung cancer tissues or adjacent normal lung tissues, the BWF samples may share more similarity with the surgical tissue samples at phylum level.

We found that *Streptococcus,* which belongs to phylum *Firmicutes,* was significantly higher in sputa than that in BWF samples and had the greatest influence on distinguishing the difference between sputum samples and BWF samples. In addition, *Streptococcus* in peripheral lung adenocarcinoma was significantly higher than central lung adenocarcinoma in the sputum group. We speculated that the differences between sputum and BWF samples was partly attributed by bacteria breed. Considering that the sputa were stored at room temperature for a period of time (< 3 h) before cryopreservation, while BWF samples were immediately transferred to cryopreservation. We need to verify whether storage at room temperature will alter taxonomy structure. However, 2 studies of pulmonary cystic fibrosis patients found that the sputum storage time at room temperature did not significantly affect the relative abundance of major pathogens in sputum samples [24, 30]. Surgery is not indicated for most patients with advanced lung cancer. Therefore, BWF samples and sputum samples are common sample types in the study of advanced lung cancer patients. We consider that BWF sample is a better substitute for lung cancer tissue compared with spontaneous sputum.

Differential genera associated with histologic types were different among different distant metastatic states of lung cancer

Previous two studies [11, 13] demonstrated differential genera between early stage lung squamous cell carcinoma and early stage lung adenocarcinoma. Since heterogeneity existed in different stages of non-small cell lung cancer [23, 31], we separately compared the lung microbiome between SCC_M0 and AD_M0, and microbiome between SCC_M1 and AD_M1.

Our study found that α diversity and β diversity of patients with different histologic types were similar, which was consistent with the conclusion of K. Leigh Greathouse et al.'s study [13]. In stage I-III, phylum *Firmicutes* and genera *Veillonella, Megasphaera, Actinomyces, Arthrobacter* were 5 differential taxonomy with mean relative abundance ≥ 0.1% that were significantly increased in AD_M0 patients. While in stage IV, genus *Capnocytophaga* and *Rothia* were 2 differential genera with mean relative abundance ≥ 0.1% that were significantly decreased in AD_M1. There were no difference in these 6 genera between former smoking group and non-smoking group, which indicated that smoking did not contribute to the alteration of the differential genera. In line with our study, previous studies comparing the lung cancer patients with different smoking status [11, 13, 15] and healthy people with different smoking status [32] also did not found significant change of the 6 genera mentioned above.

Genera *Veillonella, Megasphaera* and *Actinomyces* are oral obligate anaerobic bacterium [33-35], which were isolated and cultured from lung cancer patients [36]. Veillonella [10, 14] and Megasphaera [14] were found to be significantly increased in lung cancer patients in some clinical studies. A study which analyzed different metabolic patterns between early stage lung adenocarcinoma and early stage lung squamous cell carcinoma, reported that the ration of glucose transporter 1 (GLUT1): monocarboxylate transporter 4 (MCT4) was < 1 in adenocarcinoma and > 1 in squamous cell carcinoma [37]. Tumor cells mainly rely on anaerobic glycolysis with production of ATP and lactate even when oxygen is present [38]. Upregulation of GLUT1 supplies enough glucose into cancer cell [39] and MCT4 is responsible for transporting lactate/H+ out of the cancer cell to regulate PH level [40]. Thus, the study suggested that more lactate was transported outside the adenocarcinoma cells, which could explain the increased *Veillonella*, *Megasphaera* and *Actinomyces* in AD_M0. Our study found a significant increase of *Rothia* in SCC_M1 compared with SCC_M0. *Rothia* caused infection among immunosuppressive patients [41]. It was found that genus *Rothia* was related with the severity of COPD, and *Rothia* in sputa of GOLD C was higher than that in GOLD A and B [42]. COPD is a risk factor for lung cancer. A study has found that adenocarcinoma was the main histologic type in patients without COPD or with mild COPD, while squa-

mous cell carcinoma is the main histologic type in patients with moderate or severe COPD [43]. Moreover, the stage of tumor was positively correlated with the severity of COPD [43]. Therefore, genus *Rothia* might link severe COPD and stage IV lung squamous cell cancer. *Capnocytophaga* was another genus that was more enriched in SCC_M1. *Capnocytophag*a is a common oral bacteria, but it can also cause periodontitis [44], meningitis [45], lung abscess [46], acute exacerbation of chronic obstructive pulmonary disease [47]. Genus *Capnocytophaga* in salivary samples of lung cancer patients were significantly higher than that of healthy people [48]. Besides, *Capnocytophaga gingivalis* was evaluated in patients with oral squamous cell carcinoma (OSCC) [49]. Oral bacterial stimulation of chronic inflammation was one possible mechanism that contributed to the carcinogenesis of OSCC [49]. M. Perera et al. found that enrichment of genus Capnocytophaga was associated with LPS biosynthesis [50]. LPS are potent inflammatory molecules with cancer-promoting properties in vivo and were shown to enhance invasion in pancreatic cancer via the TLR/MyD88/NF-NF-κB pathway [51]. *Capnocytophaga* was one of the pathogens that can cause respiratory infection [46, 47]. On the analogy of the relationship between *Capnocytophaga* and OSCC, the chronic respiratory infection of *Capnocytophaga* might promote development of lung cancer, especially lung squamous cell carcinoma.

To sum up, our results suggested that differential genera between squamous cell carcinoma and adenocarcinoma existed and was related to distant metastasis states of lung cancer. We speculated that with the progress of lung cancer, either the changes of biological pattern of lung cancer would alter lung microbiome, or different bacterium played different roles in tumorgenesis of different subtypes of lung cancer.

Differential genera associated with distant metastasis were different among different histologic types of lung cancer

Lung microbiome may be associated with lung cancer metastasis [11]. Considering squamous cell carcinoma and adenocarcinoma had different distant metastasis mechanisms [20-22], we separately compared microbial characterization between AD_M0 and AD_M1, and microbiome between SCC_M0 and SCC_M1.

We found that the α diversity and the β diversity of distant metastatic lung cancer and early or locally advanced stage lung cancer were similar. In patients with adenocarcinoma, we found that phylum *Firmicutes* and genus *Streptococcus* were significantly increased in AD_M0, compared with AD_M1. Genus *Streptococcus* could predict distant metastasis of adenocarcinoma. In patients with squamous cell carcinoma, genera *Veillonella* and *Rothia* were 2 differential genera with mean relative abundance ≥ 0.1% that were significantly increased in SCC_M1, compared with SCC_M0. Genus *Veillonella* and *Rothia* chould serve as biomarkers in predicting distant metastasis of squamous cell carcinoma.

Streptococcus was decreased in AD_M1, which indicated that *Streptococcus* served as a protective role in lung adenocarcinoma. However, previous study identified that *Streptococcus* up-regulated the expression of IL-6 through Toll-like receptor 2, which enhanced matrix adhesion of lung adenocarcinoma cells and increased hepatic metastasis, which was inconsistent with our study [16]. Translocation into lung cancer tissue might be the precondition for the interaction between some bacterium and lung cancer. A study based on a metastatic lung cancer rat model reported that exposure of cigarette smoke and *Haemophilus influenzae* caused dysfunction of epithelial barrier and translocation of bacteria into tumor tissues, which synergistically promoting metastatic growth [52]. It was plausible that only when *Streptococcus* translocated into lung tissue, could it promote metastasis of lung adenocarcinoma. Thus the *Streptococcus* in the surface of lung cancer detected by BW decreased. Further researches are needed to explore the role of *Streptococcus* on distant metastasis of lung adenocarcinoma. As mentioned above, *Veillonella* is common oral bacterium and is obligate anaerobe [33, 34]. *Veillonella* may be the passengers or promoters during metastasis of squamous cell carcinoma. One study found that VEGF-B leads to neovascularization of malignant tumor, which in turn led to hypoxia in the tumor microenvironment and induces tumor metastasis; further investigation had found that the higher expression of VEGF-B

predicts poor prognosis in patients with lung squamous cell carcinoma [53]. This suggested that stage IV lung squamous cell carcinoma was more hypoxia than stage I-III, which might lead to the flourish of obligate anaerobe such as *Veillonella*. This supported the passenger hypothesis. However, Jun-Chieh J. Tsay et al. found that *Veillone*lla can activate tumor PI3K signaling pathway [10], and the aberrant PI3K was more common in lung squamous cell carcinoma than in lung adenocarcinoma [54] and was associated with distant metastasis of lung squamous cell carcinoma [55]. This suggested that *Veillonella* may play a role in promoting the metastasis of lung squamous cell carcinoma. *Rothia* was another genus that was elevated in SCC_M. As mentioned above, it was found that *Rothia* was positively related with the severity of COPD [42]. In the inflammatory environment of COPD, cancer cells enhanced their metastatic potency by epithelial- mesencymal transition [56]. Besides, several evidences suggested that *Haemophilus influenzae,* which was one of the common pathogens found in COPD and AECOPD patients, can promote metastatic lung cancer growth through regulating inflammatory mediators [52, 57]. Thus, *Rothia* might link sever COPD and distant metastasis of squamous cell carcinoma. We speculated that inhibition of genus Rothia might decrease distant metastasis potency through regulating inflammatory mediator. Interestinly, inhibition of genus Rothia may have a negative impact on cancer treatment, especially immune therapy. Lung microbiome can affect lung immune system. TIM-3, together with PD-1 have been described as hallmarks of dysfunction T cell [58]. A recent study found that COPD severity was positively correlated with the coexpression of PD-1/TIM-3 by CD8 T cells [59]. Furthermore, the study found that NSCLC patients treated by an anti-PD-1 antibody showed longer progression free survival in COPD+ patients, suggesting a higher sensitivity to PD-1 blockade in patients with COPD [59]. Another retrospective study found that a group of NSCLC patients who received antibiotics prior to immune therapy had worse clinical outcome, suggesting that antibiotic used might be a negative prognostic factor [60]. Among the study, many patients were administrated with β-lactams, which can kill genus Rothia. Therefore, we speculated that inhibition of genus Rothia might negatively impact immune therapy through regulating immune system.

Taken together, our study showed that the differential genera of distant metastatic squamous cell carcinoma were not the same as that of distant metastatic lung adenocarcinoma. This suggested that either different genera might have different mechanisms in distant metastasis of different histologic types of cancer, or the change of molecular biological characteristics during metastasis of different pathological types of cancer altered lung microbiome.

In conclusion, our study suggested that BWF samples might better reflect the lung microbiome of lung cancer tissues than sputum samples. Complex interaction existed between lung microbiome and histologic type and distant metastatic state of lung cancer. In different metastatic states of lung cancer, differential genera between squamous cell carcinoma and adenocarcinoma are different. And in different histologic types of lung cancer, distant metastasis-related genera are not the same.

Acknowledgements

This work was supported by National Natural Science foundation of China (No.81470228; 81670026; 81500023).

Disclosure of conflict of interest

None.

Abbreviations

BWF, bronchial washing fluid; SCC_M0, lung squamous cell carcinoma without distant metastasis; SCC_M1, lung squamous cell carcinoma with distant metastasis; AD_M0, lung adenocarcinoma without distant metastasis; AD_M1, lung adenocarcinoma with distant metastasis; COPD, chronic obstructive pulmonary disease; BWF_NA, lung cancer patients in BWF group without antibiotics use within 1 month; Sputum_NA, lung cancer patients in sputum group without antibiotics use within 1 month.

Address correspondence to: Drs. Hangming Dong and Shaoxi Cai, Chronic Airways Diseases Laboratory, Department of Respiratory and Critical Care Medicine, Nanfang Hospital, Southern Medical University, Guangzhou 510515, Guangdong, China. Tel: 18688426705; E-mail: huxidhm@126.com

(HMD); Tel: 13119545239; E-mail: caishaox@smu.edu.cn (SXC)

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018; 68: 7-30.

[2] Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell 2010; 140: 883-99.

[3] Hong S, Mok Y, Jeon C, Jee SH, Samet JM. Tuberculosis, smoking and risk for lung cancer incidence and mortality. Int J Cancer 2016; 139: 2447-55.

[4] Zhan P, Suo LJ, Qian Q, Shen XK, Qiu LX, Yu LK, Song Y. Chlamydia pneumoniae infection and lung cancer risk: a meta-analysis. Eur J Cancer 2011; 47: 742-7.

[5] Lin TY, Huang WY, Lin JC, Lin CL, Sung FC, Kao CH, Yeh JJ. Increased lung cancer risk among patients with pneumococcal pneumonia: a nationwide population-based cohort study. Lung 2014; 192: 159-65.

[6] Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, Moffatt MF, Cookson WO. Disordered microbial communities in asthmatic airways. PLoS One 2010; 5: e8578.

[7] Dickson RP, Erb-Downward JR, Huffnagle GB. The role of the bacterial microbiome in lung disease. Expert Rev Respir Med 2013; 7: 245-57.

[8] Cameron S, Lewis KE, Huws SA, Hegarty MJ, Lewis PD, Pachebat JA, Mur LAJ. A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. PLoS One 2017; 12: e177062.

[9] Apopa PL, Alley L, Penney RB, Arnaoutakis K, Steliga MA, Jeffus S, Bircan E, Gopalan B, Jin J, Patumcharoenpol P, Jenjaroenpun P, Wongsurawat T, Shah N, Boysen G, Ussery D, Nookaew I, Fagan P, Bebek G, Orloff MS. PARP1 is up-regulated in non-small cell lung cancer tissues in the presence of the cyanobacterial toxin microcystin. Front Microbiol 2018; 9: 1757.

[10] Tsay JJ, Wu BG, Badri MH, Clemente JC, Shen N, Meyn P, Li Y, Yie TA, Lhakhang T, Olsen E, Murthy V, Michaud G, Sulaiman I, Tsirigos A, Heguy A, Pass H, Weiden MD, Rom WN, Sterman DH, Bonneau R, Blaser MJ, Segal LN. Airway microbiota is associated with up-regulation of the PI3K pathway in lung cancer. Am J Respir Crit Care Med 2018; 198: 1188-1198.

[11] Yu G, Gail MH, Consonni D, Carugno M, Humphrys M, Pesatori AC, Caporaso NE, Goedert JJ, Ravel J, Landi MT. Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. Genome Biol 2016; 17: 163.

[12] Hosgood HD, Sapkota AR, Rothman N, Rohan T, Hu W, Xu J, Vermeulen R, He X, White JR, Wu G, Wei F, Mongodin EF, Lan Q. The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. Environ Mol Mutagen 2014; 55: 643-51.

[13] Greathouse KL, White JR, Vargas AJ, Bliskovsky VV, Beck JA, von Muhlinen N, Polley EC, Bowman ED, Khan MA, Robles AI, Cooks T, Ryan BM, Padgett N, Dzutsev AH, Trinchieri G, Pineda MA, Bilke S, Meltzer PS, Hokenstad AN, Stickrod TM, Walther-Antonio MR, Earl JP, Mell JC, Krol JE, Balashov SV, Bhat AS, Ehrlich GD, Valm A, Deming C, Conlan S, Oh J, Segre JA, Harris CC. Interaction between the microbiome and TP53 in human lung cancer. Genome Biol 2018; 19: 123.

[14] Lee SH, Sung JY, Yong D, Chun J, Kim SY, Song JH, Chung KS, Kim EY, Jung JY, Kang YA, Kim YS, Kim SK. Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. Lung Cancer 2016; 102: 89-95.

[15] Liu HX, Tao LL, Zhang J, Zhu YG, Zheng Y, Liu D, Zhou M, Ke H, Shi MM, Qu JM. Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects. Int J Cancer 2018; 142: 769-78.

[16] Gowing SD, Chow SC, Cools-Lartigue JJ, Chen CB, Najmeh S, Jiang HY, Bourdeau F, Beauchamp A, Mancini U, Angers I, Giannias B, Spicer JD, Rousseau S, Qureshi ST, Ferri LE. Gram-positive pneumonia augments non-small cell lung cancer metastasis via host toll-like receptor 2 activation. Int J Cancer 2017; 141: 561-71.

[17] Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, Sunpaweravong P, Han B, Margono B, Ichinose Y, Nishiwaki Y, Ohe Y, Yang JJ, Chewaskulyong B, Jiang H, Duffield EL, Watkins CL, Armour AA, Fukuoka M. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009; 361: 947-57.

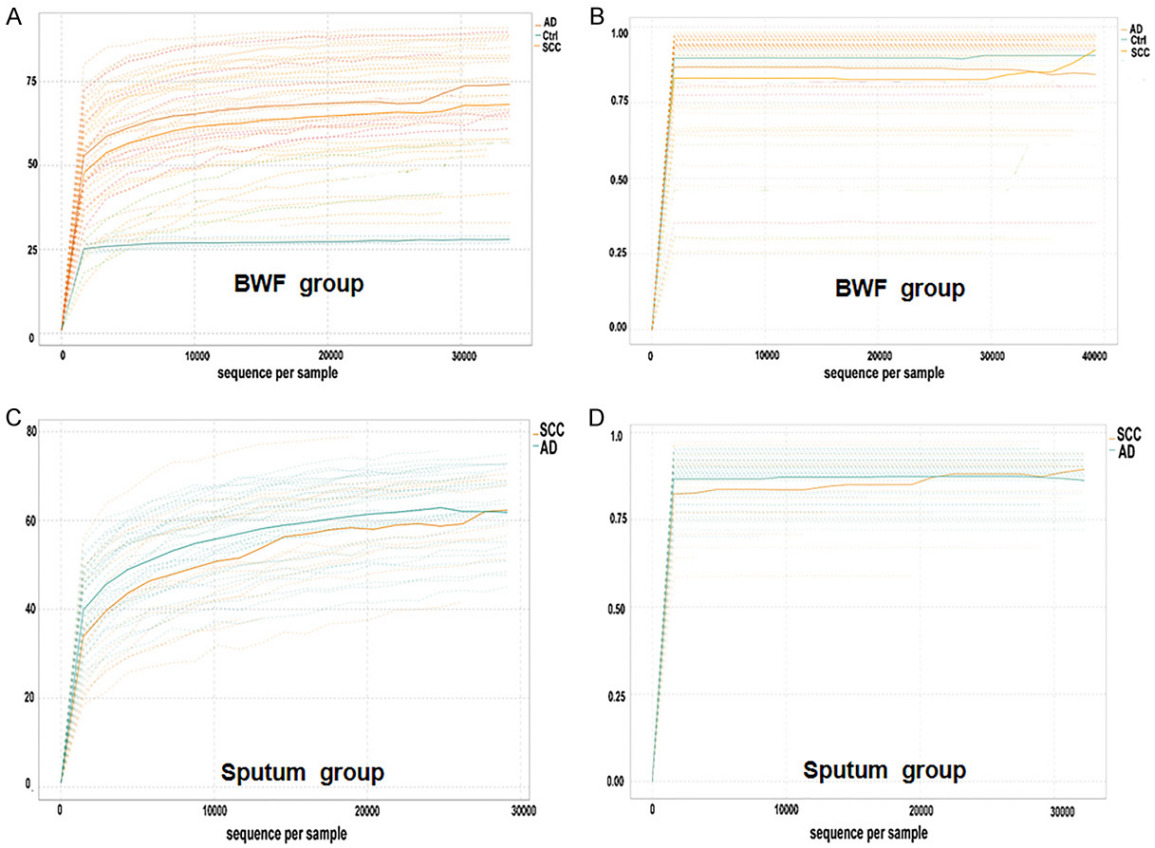
[18] Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, Powell CA, Beer D, Riely G, Garg K, Austin JH, Rusch VW, Hirsch FR, Jett J, Yang PC, Gould M; American Thoracic Society. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol 2011; 6: 244-85.

[19] Herbst RS, Heymach JV, Lippman SM. Lung cancer. N Engl J Med 2008; 359: 1367-80.

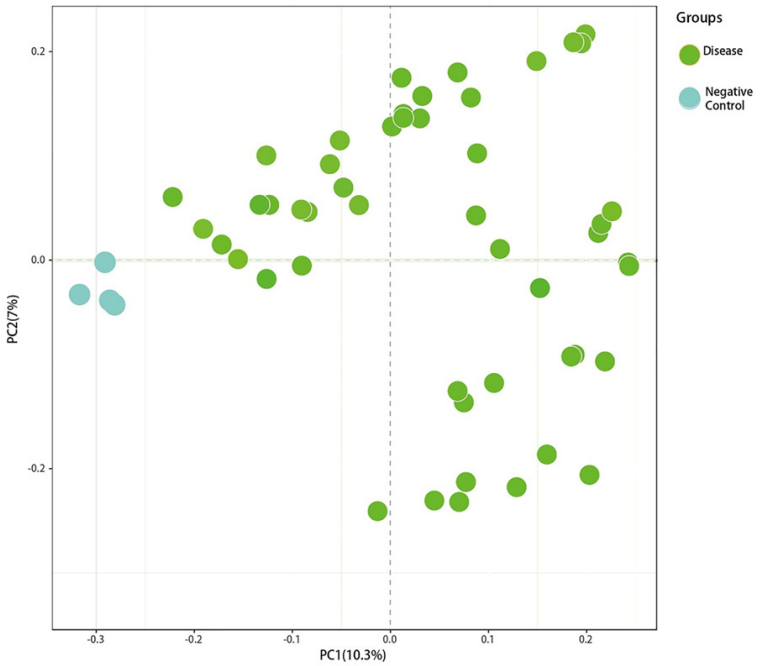
[20] Fabian K, Nemeth Z, Furak J, Tiszlavicz L, Papay J, Krenacs T, Tímár J, Moldvay J. Protein expression differences between lung adeno-

carcinoma and squamous cell carcinoma with brain metastasis. Anticancer Res 2014; 34: 5593-7.
[21] Zhong K, Yang F, Han Q, Chen J, Wang J. Skp2 expression has different clinicopathological and prognostic implications in lung adenocarcinoma and squamous cell carcinoma. Oncol Lett 2018; 16: 2873-80.
[22] Han F, Zhang MQ, Liu WB, Sun L, Hao XL, Yin L, Jiang X, Cao J, Liu JY. SOX30 specially prevents Wnt-signaling to suppress metastasis and improve prognosis of lung adenocarcinoma patients. Respir Res 2018; 19: 241.
[23] Wang J, Song J, Gao Z, Huo X, Zhang Y, Wang W, Qi J, Zheng S. Analysis of gene expression profiles of non-small cell lung cancer at different stages reveals significantly altered biological functions and candidate genes. Oncol Rep 2017; 37: 1736-46.
[24] Goddard AF, Staudinger BJ, Dowd SE, Joshi-Datar A, Wolcott RD, Aitken ML, Fligner CL, Singh PK. Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. Proc Natl Acad Sci U S A 2012; 109: 13769-74.
[25] Cabrera-Rubio R, Garcia-Nunez M, Seto L, Anto JM, Moya A, Monso E, Mira A. Microbiome diversity in the bronchial tracts of patients with chronic obstructive pulmonary disease. J Clin Microbiol 2012; 50: 3562-8.
[26] DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006; 72: 5069-72.
[27] Edgar RC. Search and clustering orders of magnitude faster than blast. Bioinformatics 2010; 26: 2460-1.
[28] Tao LL, Hu BJ, He LX, Wei L, Xie HM, Wang BQ, Li HY, Chen XH, Zhou CM, Deng WW. Etiology and antimicrobial resistance of community-acquired pneumonia in adult patients in China. Chin Med J (Engl) 2012; 125: 2967-72.
[29] Gowing SD, Chow SC, Cools-Lartigue JJ, Chen CB, Najmeh S, Jiang HY, Bourdeau F, Beauchamp A, Mancini U, Angers I, Giannias B, Spicer JD, Rousseau S, Qureshi ST, Ferri LE. Gram-positive pneumonia augments non-small cell lung cancer metastasis via host toll-like receptor 2 activation. Int J Cancer 2017; 141: 561-71.
[30] Cuthbertson L, Rogers GB, Walker AW, Oliver A, Hafiz T, Hoffman LR, Carroll MP, Parkhill J, Bruce KD, van der Gast CJ. Time between collection and storage significantly influences bacterial sequence composition in sputum samples from cystic fibrosis respiratory infections. J Clin Microbiol 2014; 52: 3011-6.
[31] Babadi AS, Kiani A, Mortaz E, Taghavi K, Khosravi A, Marjani M, Seifi S, Emami H, Abedini A. Serum interleukin-27 level in different clinical stages of lung cancer. Open Access Maced J Med Sci 2019; 7: 45-9.
[32] Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, Flores SC, Fontenot AP, Ghedin E, Huang L, Jablonski K, Kleerup E, Lynch SV, Sodergren E, Twigg H, Young VB, Bassis CM, Venkataraman A, Schmidt TM, Weinstock GM; Lung HIV Microbiome Project. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med 2013; 187: 1067-75.
[33] Mur LA, Huws SA, Cameron SJ, Lewis PD, Lewis KE. Lung cancer: a new frontier for microbiome research and clinical translation. Ecancermedicalscience 2018; 12: 866.
[34] Lanjekar VB, Marathe NP, Ramana VV, Shouche YS, Ranade DR. Megasphaera indica sp. nov., an obligate anaerobic bacteria isolated from human faeces. Int J Syst Evol Microbiol 2014; 64: 2250-6.
[35] Topcuoglu N, Erdem AP, Karacan I, Kulekci G. Oral microbial dysbiosis in patients with kostmann syndrome. J Med Microbiol 2019; 68: 609-15.
[36] Rybojad P, Los R, Sawicki M, Tabarkiewicz J, Malm A. Anaerobic bacteria colonizing the lower airways in lung cancer patients. Folia Histochem Cytobiol 2011; 49: 263-6.
[37] Meijer TW, Schuurbiers OC, Kaanders JH, Looijen-Salamon MG, de Geus-Oei LF, Verhagen AF, Lok J, van der Heijden HF, Rademakers SE, Span PN, Bussink J. Differences in metabolism between adeno- and squamous cell non-small cell lung carcinomas: spatial distribution and prognostic value of GLUT1 and MCT4. Lung Cancer 2012; 76: 316-23.
[38] Busk M, Horsman MR, Jakobsen S, Bussink J, van der Kogel A, Overgaard J. Cellular uptake of PET tracers of glucose metabolism and hypoxia and their linkage. Eur J Nucl Med Mol Imaging 2008; 35: 2294-303.
[39] Macheda ML, Rogers S, Best JD. Molecular and cellular regulation of glucose transporter (GLUT) proteins in cancer. J Cell Physiol 2005; 202: 654-62.
[40] Gillies RJ, Robey I, Gatenby RA. Causes and consequences of increased glucose metabolism of cancers. J Nucl Med 2008; 49 Suppl 2: 24S-42S.
[41] Khan ST, Ahamed M, Musarrat J, Al-Khedhairy AA. Anti-biofilm and antibacterial activities of zinc oxide nanoparticles against the oral opportunistic pathogens rothia dentocariosa and rothia mucilaginosa. Eur J Oral Sci 2014; 122: 397-403.
[42] Lee SW, Kuan CS, Wu LS, Weng JT. Metagenome and metatranscriptome profiling of moderate

and severe copd sputum in Taiwanese han males. PLoS One 2016; 11: e159066.
[43] Sekine Y, Suzuki H, Yamada Y, Koh E, Yoshino I. Severity of chronic obstructive pulmonary disease and its relationship to lung cancer prognosis after surgical resection. Thorac Cardiovasc Surg 2013; 61: 124-30.
[44] Nonnenmacher C, Mutters R, de Jacoby LF. Microbiological characteristics of subgingival microbiota in adult periodontitis, localized juvenile periodontitis and rapidly progressive periodontitis subjects. Clin Microbiol Infect 2001; 7: 213-7.
[45] de Boer MG, Lambregts PC, van Dam AP, van T WJ. Meningitis caused by Capnocytophaga canimorsus: when to expect the unexpected. Clin Neurol Neurosurg 2007; 109: 393-8.
[46] Migiyama Y, Anai M, Kashiwabara K, Tomita Y, Saeki S, Nakamura K, Okamoto S, Ichiyasu H, Fujii K, Kohrogi H. Lung abscess following bronchoscopy due to multidrug-resistant capnocytophaga sputigena adjacent to lung cancer with high PD-L1 expression. J Infect Chemother 2018; 24: 852-5.
[47] Tan L, Wang H, Li C, Pan Y. 16S rDNA-based metagenomic analysis of dental plaque and lung bacteria in patients with severe acute exacerbations of chronic obstructive pulmonary disease. J Periodontal Res 2014; 49: 760-9.
[48] Yan X, Yang M, Liu J, Gao R, Hu J, Li J, Zhang L, Shi Y, Guo H, Cheng J, Razi M, Pang S, Yu X, Hu S. Discovery and validation of potential bacterial biomarkers for lung cancer. Am J Cancer Res 2015; 5: 3111-22.
[49] Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, Goodson JM. The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. J Transl Med 2005; 3: 27.
[50] Perera M, Al-Hebshi NN, Perera I, Ipe D, Ulett GC, Speicher DJ, Chen T, Johnson NW. Inflammatory bacteriome and oral squamous cell carcinoma. J Dent Res 2018; 97: 725-32.
[51] Ikebe M, Kitaura Y, Nakamura M, Tanaka H, Yamasaki A, Nagai S, Wada J, Yanai K, Koga K, Sato N, Kubo M, Tanaka M, Onishi H, Katano M. Lipopolysaccharide (LPS) increases the invasive ability of pancreatic cancer cells through the TLR4/MyD88 signaling pathway. J Surg Oncol 2009; 100: 725-31.
[52] Jungnickel C, Wonnenberg B, Karabiber O, Wolf A, Voss M, Wolf L, Honecker A, Kamyschnikow A, Herr C, Bals R, Beisswenger C. Cigarette smoke-induced disruption of pulmonary barrier and bacterial translocation drive tumor-associated inflammation and growth. Am J Physiol Lung Cell Mol Physiol 2015; 309: L605-13.
[53] Yang X, Zhang Y, Hosaka K, Andersson P, Wang J, Tholander F, Cao Z, Morikawa H, Tegnér J, Yang Y, Iwamoto H, Lim S, Cao Y. VEGF-B promotes cancer metastasis through a VEGF-A-independent mechanism and serves as a marker of poor prognosis for cancer patients. Proc Natl Acad Sci U S A 2015; 112: E2900-9.
[54] Zhang Y, Wang DC, Shi L, Zhu B, Min Z, Jin J. Genome analyses identify the genetic modification of lung cancer subtypes. Semin Cancer Biol 2017; 42: 20-30.
[55] Paik PK, Shen R, Won H, Rekhtman N, Wang L, Sima CS, Arora A, Seshan V, Ladanyi M, Berger MF, Kris MG. Next-generation sequencing of stage iv squamous cell lung cancers reveals an association of pi3k aberrations and evidence of clonal heterogeneity in patients with brain metastases. Cancer Discov 2015; 5: 610-21.
[56] Sekine Y, Hata A, Koh E, Hiroshima K. Lung carcinogenesis from chronic obstructive pulmonary disease: characteristics of lung cancer from COPD and contribution of signal transducers and lung stem cells in the inflammatory microenvironment. Gen Thorac Cardiovasc Surg 2014; 62: 415-21.
[57] Jungnickel C, Schmidt LH, Bittigkoffer L, Wolf L, Wolf A, Ritzmann F, Kamyschnikow A, Herr C, Menger MD, Spieker T, Wiewrodt R, Bals R, Beisswenger C. IL-17C mediates the recruitment of tumor-associated neutrophils and lung tumor growth. Oncogene 2017; 36: 4182-90.
[58] Thommen DS, Schreiner J, Muller P, Herzig P, Roller A, Belousov A, Umana P, Pisa P, Klein C, Bacac M, Fischer OS, Moersig W, Savic Prince S, Levitsky V, Karanikas V, Lardinois D, Zippelius A. Progression of lung cancer is associated with increased dysfunction of T cells defined by coexpression of multiple inhibitory receptors. Cancer Immunol Res 2015; 3: 1344-55.
[59] Biton J, Ouakrim H, Dechartres A, Alifano M, Mansuet-Lupo A, Si H, Halpin R, Creasy T, Bantsimba-Malanda C, Arrondeau J, Goldwasser F, Boudou-Rouquette P, Fournel L, Roche N, Burgel PR, Goc J, Devi-Marulkar P, Germain C, Dieu-Nosjean MC, Cremer I, Herbst R, Damotte D. Impaired tumor-infiltrating t cells in patients with chronic obstructive pulmonary disease impact lung cancer response to PD-1 blockade. Am J Respir Crit Care Med 2018; 198: 928-40.
[60] Zhao S, Gao G, Li W, Li X, Zhao C, Jiang T, Jia Y, He Y, Li A, Su C, Ren S, Chen X, Zhou C. Antibiotics are associated with attenuated efficacy of anti-PD-1/PD-L1 therapies in Chinese patients with advanced non-small cell lung cancer. Lung Cancer 2019; 130: 10-7.



Supplementary Figure 1. OTU and Simpson rarefaction curve of lung cancer patients. (A) OTU rarefaction curve of BWF group; (B) Simpson rarefaction curve of BWF group; (C) OTU rarefaction curve of sputum group; (D) Simpson rarefaction curve of sputum group.

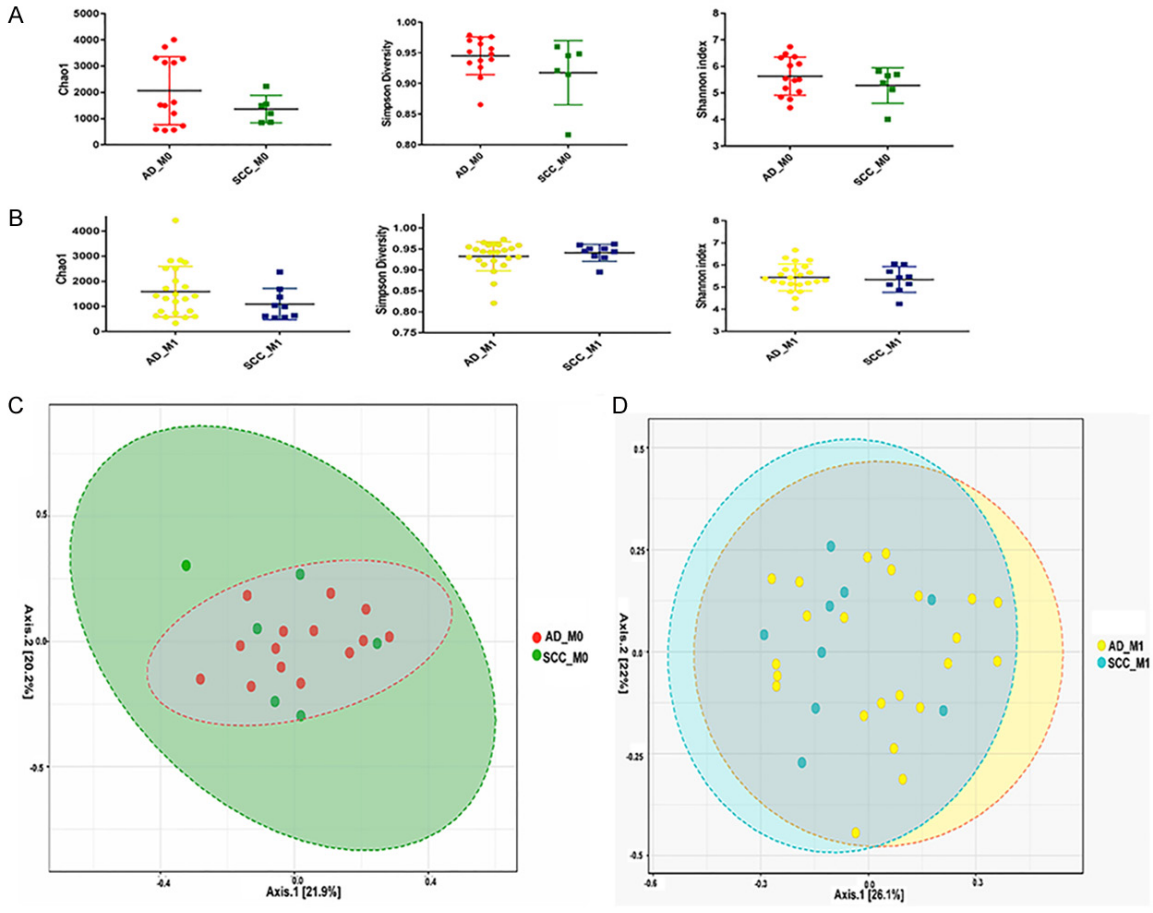


Supplementary Figure 2. Unweighted unifrac PCOA plot of BWF microbiome between Disease group and negative controls.

Supplementary Table 1. Demographics and clinical characteristics among central lung adenocarcinoma and peripheral lung adenocarcinoma of sputum group

	Central AD	Peripheral AD	*P* value
N	11	26	
Age-mean (SD)	55.18 (10.48)	58.50 (8.95)	0.334[a]
Gender			0.235[b]
Male, n (%)	4	15	
Female, n (%)	7	11	
BMI (kg/m^2)-mean (SD)	21.50 (2.66)	22.16 (4.04)	0.372[a]
Smoking Status			0.285[b]
Current or former Smoker, n (%)	3	12	
Never smoker, n (%)	8	14	
Antibiotics use within 1 month			0.8[b]
Yes, n (%)	6	13	
No, n (%)	5	13	

a: *p* values were calculated by independent t test; b: *p* values were calculated by Fisher exact test.



Supplementary Figure 3. Comparison of α diversity and β diversity of sputum microbiome from different pathological types. (A) α diversity between AD_M0 and SCC_M0; (B) α diversity between AD_M1 and SCC_M1; (C) PCOA plot based on Bray-Curtis distance of BWF microbiome between AD_M0 and SCC_M0; (D) PCOA plot based on Bray-Curtis distance of BWF microbiome between AD_M1 and SCC_M1.

Supplementary Table 2. Differential taxonomy identified by LEFse between AD_M0 and SCC_M0 of BWF group

Taxonomy	AD_M0 mean	SCC_M0 mean	*p*-value	LDA score
p__Tenericutes	0.000790677	0.004610612	0.035005682	2.669532714
p__Firmicutes	0.447776807	0.228030768	0.035005682	4.353749711
c__Thermoleophilia	0.000576772	0.000163159	0.046670224	3.081739516
c__Mollicutes	0.000783396	0.004522269	0.047645426	2.670579382
c__Erysipelotrichi	0.002085398	0.001264239	0.047645426	2.498198687
c__Clostridia	0.180491846	0.062487193	0.01271625	4.098320561
o__Solirubrobacterales	0.000377006	0.00004766	0.003648341	2.944425012
o__Erysipelotrichales	0.002085398	0.001264239	0.047645426	2.498198687
o__Clostridiales	0.180393116	0.062487193	0.01271625	4.097961952
o__Caulobacterales	0.027026373	0.005781399	0.047645426	3.390082327
f__Vibrionaceae	0.001964156	0.000105784	0.007817424	2.765867789
f__Veillonellaceae	0.154413	0.040038646	0.01271625	4.081739822
f__Oxalobacteraceae	0.00556554	0.001497878	0.047645426	2.722980782
f__Microbacteriaceae	0.001270621	0.000355287	0.024706434	2.655684682
f__Erysipelotrichaceae	0.002085398	0.001264239	0.047645426	2.498198687
f__Cellulomonadaceae	0.000105767	1.43401E-05	0.040189713	3.412521673
f__Caulobacteraceae	0.026759239	0.005748468	0.047645426	3.385872797
f__Bradyrhizobiaceae	0.00584567	0.001116444	0.018086461	2.795782191
f__Actinomycetaceae	0.00580753	0.001930303	0.035005682	2.675197843
g__Vibrio	0.000410117	2.61929E-05	0.047873278	3.045629725
g__Veillonella	0.125105038	0.033581439	0.018086461	3.987029577
g__Sediminibacterium	0.000568495	0.000380834	0.019945403	2.666305004
g__Methyloversatilis	0.000623171	0.000432561	0.034207141	2.643365534
g__Megasphaera	0.018561356	0.002802489	0.008808617	3.210978727
g__Bulleidia	0.002020059	0.001035811	0.046670224	2.473199279
g__Bradyrhizobium	0.000184427	1.64657E-05	0.028286006	3.223872605
g__Bacillus	0.001716238	0.000731994	0.046670224	2.344338777
g__Arthrobacter	0.009453996	0.000995771	0.028059661	2.998869239
g__Actinomyces	0.00580753	0.001923715	0.035005682	2.675358203

Supplementary Table 3. Differential taxonomy identified by LEFse between AD_M1 and SCC_M1 of BWF group

Taxonomy	AD_M1 mean	SCC_M1 mean	*p*-value	LDA score
p__Nitrospirae	0	2.42571E-05	0.004459129	1.841627418
c__Nitrospira	0	2.42571E-05	0.004459129	1.845952499
o__Nitrospirales	0	2.42571E-05	0.004459129	1.837337118
o__Desulfovibrionales	3.15671E-05	0.000744571	0.022331373	1.891325505
f__Streptomycetaceae	1.48885E-05	0.000185125	0.038476082	1.637121434
f__Rikenellaceae	6.92138E-05	5.31801E-05	0.043691175	1.584097684
f__Nitrospiraceae	0	1.71286E-05	0.004459129	1.960899639
f__Flavobacteriaceae	0.012185503	0.022097757	0.01424808	3.046580496
f__Enterococcaceae	0.000845131	6.95714E-06	0.028286025	2.023446077
f__Desulfovibrionaceae	3.00671E-05	0.000594718	0.028286025	1.801891168
f__Deinococcaceae	6.59167E-06	0.000196539	0.03349736	1.59542452

g__Succiniclasticum	0	2.19571E-05	0.016626246	2.245797927
g__Streptomyces	1.14635E-05	9.73744E-05	0.014639429	1.637780531
g__Scardovia	0	3.28714E-05	0.016626246	1.827142075
g__Rothia	0.000884501	0.00310511	0.014078327	2.383309592
g__Rhodobacter	9.68108E-05	0.001049386	0.03349736	2.023448313
g__Planomicrobium	0	0.000265543	0.016626246	1.867823424
g__Nitrospira	0	1.01714E-05	0.016626246	2.33313592
g__Enterococcus	0.000744207	0	0.016740658	2.02401274
g__Enhydrobacter	0.000358123	0.000018	0.03025203	1.783587458
g__Desulfovibrio	2.59921E-05	0.000330309	0.01867354	1.625188502
g__Deinococcus	6.59167E-06	0.000196539	0.03349736	1.613437849
g__Coprococcus	0.000140136	0.00023621	0.024711443	1.443343046
g__Chryseobacterium	9.5135E-06	5.13643E-05	0.03349736	1.993510188
g__Capnocytophaga	0.011049623	0.021133776	0.017960478	3.057171958
g__Blvii28	0	2.35571E-05	0.016626246	1.927621828
g__Blautia	1.79429E-05	0.000325919	0.019102172	1.648421693

Supplementary Table 4. Differential taxonomy identified by LEFse between AD_M0 and AD_M1 of BWF group

Taxonomy	AD_M0 mean	AD_M1 mean	*p*-value	LDA score
p__Firmicutes	0.447776807	0.288118957	0.022494271	4.188231899
c__Thermoleophilia	0.000576772	0.000163047	0.012793708	2.731914812
c__Bacilli	0.265199562	0.138350939	0.042522478	4.113789919
o__Solirubrobacterales	0.000377006	0.000127612	0.004954139	2.77704709
o__Lactobacillales	0.242551401	0.102358837	0.042522478	4.153844539
f__Streptococcaceae	0.227034496	0.087803271	0.042522478	4.150618597
f__Shewanellaceae	5.55387E-05	0	0.016626246	2.910297165
f__EB1017	9.00317E-05	0	0.016626246	2.910751495
f__Cellulomonadaceae	0.000105767	2.94519E-05	0.044375994	3.163104864
f__Aerococcaceae	8.05927E-05	0.001382576	0.046214074	2.314747913
g__Streptococcus	0.226358387	0.086533436	0.042522478	4.152190952
g__Shewanella	5.55387E-05	0	0.016626246	2.92702161
g__Dorea	0.000056037	0	0.016626246	2.947445556
g__Cellvibrio	8.35529E-05	0	0.016626246	2.973230491

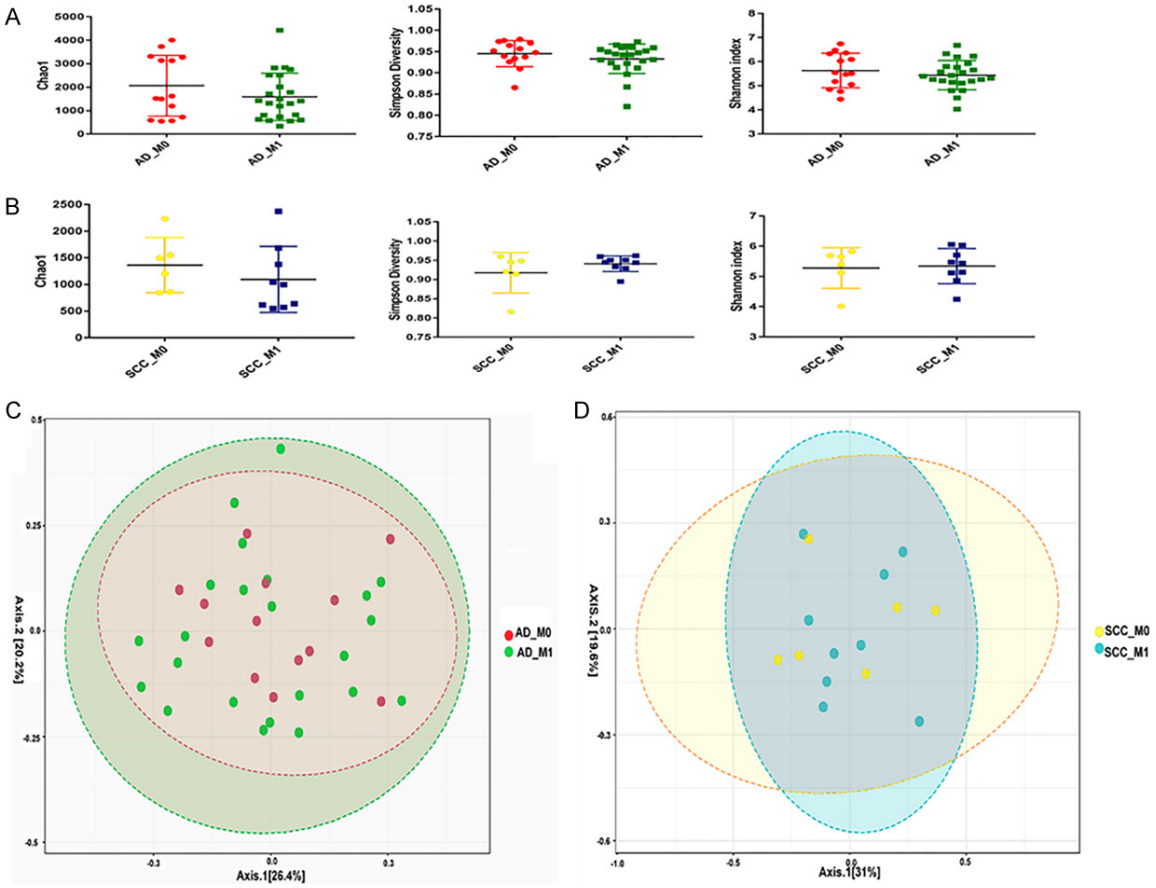
Supplementary Table 5. Differential taxonomy identified by LEFse between SCC_M0 and SCC_M1 of BWF group

Taxonomy	SCC_M0 mean	SCC_M1 mean	*p*-value	LDA score
c__Cytophagia	0.000667624	5.96201E-05	0.023766877	2.120447791
o__Cytophagales	0.000667624	5.96201E-05	0.023766877	2.12861783
o__Bifidobacteriales	0	8.25893E-05	0.025046192	2.735963673
f__Veillonellaceae	0.040038646	0.126607991	0.025347319	3.90901044
f__Desulfovibrionaceae	1.08571E-05	0.000594718	0.034435123	2.077111415
f__Bifidobacteriaceae	0	8.25893E-05	0.025046192	2.733983871
g__Veillonella	0.033581439	0.110334253	0.01271625	3.854586826
g__Ruminococcus	0	8.16883E-05	0.025046192	2.776335103

g__Rothia	0.000671144	0.00310511	0.017959832	2.453780857
g__Pyramidobacter	1.45857E-05	0.000278008	0.047873278	2.551506115
g__Erwinia	0.000306273	0	0.025046192	2.306777287
g__Desulfovibrio	1.08571E-05	0.000330309	0.034435123	2.692273793
g__Blautia	0	0.000325919	0.009161627	2.251000804

Supplementary Table 6. Differential taxonomy identified by LEFse between current or former smokers and non-smokers of BWF group

Taxonomy	Current or former smokers mean	Non-smoker mean	*p*-value	LDA score
c__VC2_1_Bac22	0	3.17134E-05	0.042254021	0.810659242
c__SC3	0	4.20875E-05	0.042254021	0.813536579
c__Nostocophycideae	0	0.000216468	0.042254021	1.534917566
c__ML635J_21	0	1.90358E-05	0.042254021	0.84693079
o__Syntrophobacterales	0	5.86592E-05	0.042254021	0.674972936
o__Nostocales	0.000131275	5.63636E-06	0.042254021	1.007050002
o__Legionellales	0	1.90358E-05	0.045219245	1.19658592
o__11_24	0	9.43636E-06	0.042254021	1.004800747
f__Syntrophobacteraceae	0	4.16727E-05	0.042254021	0.731568137
f__Haliangiaceae	0.000398235	0	0.042254021	0.547618106
f__Glycomycetaceae	0	3.44545E-06	0.042254021	0.85874035
g__Ruminococcus	0	9.12727E-06	0.039742759	1.705174108
g__Propionivibrio	0	0.000248912	0.042254021	0.555000762
g__Parabacteroides	0	5.25029E-05	0.042254021	0.588050521
g__Nocardia	0	4.16727E-05	0.042254021	1.467746125
g__Methylotenera	0	3.17134E-05	0.042254021	0.960048178
g__Glycomyces	0	4.20875E-05	0.042254021	0.88534792



Supplementary Figure 4. Comparison of α diversity and β diversity of sputum microbiome from different metastatic states. (A) α diversity between AD_M0 and AD_M1; (B) α diversity between SCC_M0 and SCC_M1; (C) PCOA plot based on Bray-Curtis distance of BWF microbiome between AD_M0 and AD_M1; (D) PCOA plot based on Bray-Curtis distance of BWF microbiome between SCC_M0 and SCC_M1.